In the Specification:

Change the paragraph that begins on page 1, line 9 to:

A speech recognizer trained with relatively a quiet office environment speech data and then operating in a mobile environment may fail due to at least to the ~~tow~~ two distortion sources of back ground noise and microphone changes. The background noise may, for example, be from a computer fan, car engine, and/or road noise. The microphone ~~changer~~ changes may be due to the quality of the microphone, whether the microphone is hand-held or hands-free and, ~~a~~ the position of the microphone to the mouth. In mobile applications of speech recognition, both the microphone conditions~~er~~ and background noise are subject to change.

Change the paragraph that begins on page 1, line 17 and continues to the end of the page to:

Cepstral Mean Normalization (CMN) removes utterance mean and is a simple and effective way of dealing with convolutive distortion such as telephone channel distortion. See "Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification" of B. Atal in Journal of Acoustics Society of America, Vol. 55: 1304–1312, 1974. Spectral Subtraction (SS) reduces background noise in the feature space. See article "Suppression of Acoustic Noise in Speech Using Spectral Subtraction" of S.F. Boll in IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-27(2): 113-129, April 1979. Parallel Model Combination (PMC) gives an approximation of speech models in noisy conditions from noise-free speech models and noise estimates. See "An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise" of M.J. F. ~~Glaes~~ Gales and S. Young in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Volume 1, pages 233-236, U.S.A., April 1992. The techniques do not require any training data.

Change the paragraph that begins at the top of page 2, line 2 to line 21 to:

Joint compensation of ~~additional~~ additive noise and convolutive noise can be achieved by the introduction of a channel model and a noise model. A spectral bias for additive noise and a cepstral bias for convolutive noise are introduced in an article by M. Afify, Y. Gong, and J. P. Haton. This article is entitled "A General Joint Additive and Convolutive Bias Compensation Approach Applied to Noisy Lombard Speech Recognition" in IEEE Trans. on Speech and Audio Processing, 6(6): 524-538, November 1998. The ~~five~~ two biases can be calculated by application of Expectation Maximization (EM) in both spectral and convolutive domains. A procedure by J.L. Gauvain, et al, is presented to calculate the convolutive component, which requires rescanning of training data. See J.L. Gauvain, L. Lamel, M. Adda-Decker, and D. Matrouf entitled "Developments in Continuous Speech Dictation using the ARPA NAB News Task." In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 73-76, Detroit, 1996. Solution of the convolutive component by a steepest descent method has also been reported. See Y. Minami and S. Furui entitled "A Maximum Likelihood Procedure for a Universal Adaptation Method Based on HMM Composition." See Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, pages 129-132, Detroit, 1995. A method by Y. Minami and S. Furui needs additional universal speech models, and ~~redestination~~ re-estimation of channel distortion with the universal models when channel changes. See Y. Minami and S. Furui entitled "Adaptation Method Based on HMM Composition and EM Algorithm" in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pages 327-330, Atlanta 1996.

Change the paragraph that begins on page 2, line 28 and continues through page 3, line 5 to:

Alternatively, the nonlinear changes of both type of distortions can be approximated by linear equations, assuming that the changes are small. A Jacobian approach, which models speech model parameter changes as the product of a jacobian matrix and the difference in noisy conditions, and statistical linear approximation are along this direction. See S. Sagayama, Y. Yamaguchi, and S. Takahashi entitled "Jacobian Adaptation of Noisy Speech Models," in Proceedings of IEEE Automatic

Speech Recognition Workshop, pages 396-403, Santa Barbara, CA, USA, December 1997. IEEE Signal Processing Society. Also see "Statistical Linear Approximation for Environment Compensation" of N.S. Kim, IEEE Signal Processing Letters, 5(1): 8-10, January 1998.

Change the paragraph that begins on page 3, line 15 as follows:

In accordance with one establishment of the present ~~inventor~~ invention a new method is disclosed that ~~Handles~~ simultaneously handles noise and channel distortions to make a speaker independent system robust to a wide variety of noises and channel distortions.

Change the paragraph on page 3, line 25 to:

Figure 2 illustrates the method of the present invention. ~~Generating~~

Change the paragraph that begins on page 4, line 3 to:

Referring to Fig. 1 there is illustrated a speech recognizer according to the present invention. The speech is applied to recognizer 11. The speech is compared to Hidden Markov Models (HMM) 13 to recognize the text. The models initially provided ~~on~~ are those ~~with~~ based on speech recorded in a quiet environment and ~~the~~ with a microphone of good quality. We want to develop a speech model set suitable for operating in the simultaneous presence of channel/microphone distortion and background noise. In accordance with the present invention, a speech model set is provided using statistics about the noise and speech~~. A low computation cost method integrates both PMC and CMN.~~

Change the paragraph that begins on page 4, line 12 as follows:

Referring to Figure 2, the first Step 1 is to start with HMM models trained on clean speech, with cepstral mean normalization. We modify these models to get models to compensate for channel/microphone distortion (convolutive distortion) and

simultaneous background noise (additive distortion). The HMM modeling method of this invention represents the acoustic probability density function (PDF) corresponding to each HMM state as a mixture of Gaussian components, as is well known in the art. ~~For an~~ Such HMM models~~, we~~ have ~~a lot of~~ many parameters, such as Gaussian component mean vectors, covariances, and mixture component weights for each state, as well as HMM state transition probabilities. The method of this invention teaches modifying ~~but only change one subset of the parameters and that is~~ the mean vectors $m_{p,j,k}$~~.~~ of the original model space, ~~The mean vectors $m_{p,j,k}$ of the original model space is modified~~ where p is the index of the ~~Probability Density Function (PDF)~~ HMM, j is the state and k is the mixing component.

Change the paragraph that begins on page 4, line 20 as follows:

The second Step 2 is to calculate ~~which is~~ the mean mel-scaled cesptrum coefficients (MFCC) vector over the trained database. Scan all data and calculate the mean to get ~~$\hat{b}$~~ b.

Change the paragraph that begins on page 4, line 23 as follows:

The third Step 3 is to add mean ~~$\hat{b}$~~ b to each of this mean vector pool represented by $m_{p,j,k}$ equation (1) to get:

$$\overline{m}_{p,j,k} = m_{p,j,k} + b. \qquad (1)$$

Change the paragraph that begins on page 4, line 28 to:
For example, there could be 100 ~~PDF~~ HMMs, 3 states per ~~PDF~~ HMM and 2 vectors per state, or a total of 600 vectors.

Change the paragraph that begins on page 5, line 15 to:

In Step 5, we calculate the mean vectors adapted to the noise $\tilde{X}$ using equation 4.

$$\hat{m}_{p,j,k} = IDFT(DFT(\overline{m}_{p,j,k}) \oplus DFT(\tilde{X})). \qquad (4)$$

where *DFT* and *IDFT* are, respectively, the DFT and inverse DFT operation, ~~$\bar{m}_{p,j,k}$~~ $\hat{m}_{p,j,k}$ is the noise compensated mean vector.

Change the paragraph that begins on page 5, line 21 and ends on page 6, line 2 to:

Equation 4 involves several operators. DFT is the Discrete Fourier Transform and IDFT is the Inverse Discrete Fourier Transform, which are respectively used to convert from the cepstrum domain to the log spectrum domain, and vice versa. The $\oplus$ is an operation applied to two log spectral vectors to produce a log spectral vector representing the linear sum of spectra. ~~with two vectors. A $\oplus$ B = C. How~~ The operation $\oplus$ is defined~~, we look at~~ by equations 2 and 3. Equation 2 ~~says~~ defines the operation [[+]] $\oplus$ which operates on two D dimensional vectors u and v and the result is a vector of D dimensions, ~~or~~ $[w_1', w_2', \cdots w_D']^T$ where T is the transposition. ~~We take the two vectors and produce another vector. We need to specify each element in the resultant vector.~~ Equation 3 defines ~~says that~~ the jth element in that vector (w_j') ~~is defined by the exponential of the element of u added to the exponential if the jth element of v and take the log of the combination of the exponential of u added to the exponential of the j the element of v~~. This completes the definition of Equation 4.

Change the paragraph that begins on page 6, line 4 to:

In the following steps, we need to remove the mean vector $\hat{b}$ of the noisy data y over the noisy speech space $\mathcal{N}$ (from the resultant model). One may be able to synthesize enough noisy data from compensated models but this requires a lot of calculation. In accordance with the present invention the vector is calculated using statistics of the noisy models. The whole recognizer will operate with CMN (cepstral mean normalization mode), but the models in Equation 4 are no longer mean normalized. We have dealt with additive noise. The second half of the processing is removing the cepstral mean of our models defined in Equation 4. This is not difficult because we have the models in Equation 4. In Step 6, we need to integrate all the samples generated by Equation 4 to get the mean~~. Mean is~~ $\hat{b}$. Equation 5 is this integration.

Change the paragraph that begins on page 6, line 14 to:

Let $\mathcal{H}$ be the variable denoting ~~PDF~~ HMM index, J be the variable for state index, and $\mathcal{K}$ be the variable for mixing component index.

$$\hat{b} = E\{y\} \tag{5}$$

$$= \int_H y \sum_p \sum_j \sum_k P_{\mathcal{H}}(p) P_{J|\mathcal{H}}(j|p) P_{\mathcal{K}|\mathcal{H}J}(k|p,j) P_{\mathbf{Y}|\mathcal{H}J,\mathcal{K}}(\mathbf{y}|p,j,k) d\mathbf{y}$$

Change the paragraph that begins on page 7, line 1 to:

Equation 7 shows that $\hat{b}$ can be worked out analytically, and it is not necessary to do the physical~~ly~~ generation and integration. The final result is represented by Equation 7 which reduces ~~is~~ the integration into ~~several~~ sums~~. Sums~~ over HMMs, ~~probability density functions and the sum,~~ over states and ~~sum~~ over mixing components. ~~Then you have several quantities. The P_H is the probability of having the PDF index. The P_J given $\mathcal{H}$ is the probability of being in the state if given the PDFp. The next is the probability the mixing component p,j given we have the PDF index. The mean vector of the compensated mode. To make this complete~~ Finally the estimated noise-compensated channel bias, ~~we remove this~~ $\hat{b}$, is removed from the compensated model means to get the target model means. This is Step 7. The target model is:

$$\dot{m}_{p,j,k} = \hat{m}_{p,j,k} - \hat{b} \tag{8}$$

Change the paragraph that begins on page 7, line 14 to:

This resulting target model means are the desired modified parameters of the HMM models used in the ~~is what we want to load into our~~ recognizer. This operation is done

for each utterance. Figure 2 illustrates that for a next utterance (Step 8) the process starts with step 4.

Change the paragraph that begins on page 7, line 16 to:

Calculation of $\hat{b}$ ~~thus~~ requires the knowledge of the probabilit~~y~~ies of each PDF. There are two issues with ~~P ($\mathcal{H}$= p):~~ the probabilities:

- ~~It~~ They needs additional storage space.
- ~~It is~~ They are dependent of the recognition task e.g. vocabulary, grammar.

Change the paragraph beginning on page 7, line 22 to:

Although it is possible to obtain ~~that~~ the probabilit~~y~~ies, we ~~want to~~ can also consider the following simplified case~~s~~.

Change the paragraph beginning on page 7, line 24 and continuing to page 8, line15:

~~This~~ The operation~~s~~ to calculate ~~the~~ $\hat{b}$ can be simplified by assuming ~~with three approximations. The first one uses equal probabilities for $P_{\mathcal{H}}(p)$ or constraint C.~~

~~1. Use equal probabilities for $P_{\mathcal{H}}(p)$:~~

~~$P_{\mathcal{H}}(p) = C$~~ ~~(9)~~

~~2. Use~~ equal probabilities for $P_{\mathcal{H}}(p)$, $P_{J|\mathcal{H}}P(j|p)$ and $P_{K|\mathcal{H},J}(k|p,j)$.

$$P_{\mathcal{H}}(p) = C$$
$$P_{J|\mathcal{H}}(j|p) = D \qquad (10)$$
$$P_{K|\mathcal{H},J}(k|p,j) = E$$

C, D and E are selected such that they represent equal probabilities. Therefore we have the following: C is chosen such that it provides a probability such that each HMM is likely, so C=1/(number of HMM models); D is chosen such that each state of a given HMM is equally likely, where the HMM is indexed by p, so D=1/(number of states in HMM(p)); and E is chosen such that each mixing component of a state of an HMM is equally likely, where the state of an HMM is indexed by j, so E=1/(number of mixing components in HMM(p) state(j).

~~3.~~ In fact, the case described in Eq-10 consists in averaging the compensated mean vectors ~~$\overline{m}$~~ $\hat{m}_{p,j,k}$. Referring to Eq-4 and Eq-1, it can be expected that the averaging reduces the speech part $m_{p,j,k}$ just as CMN does. Therefore, Eq-7 could be further simplified into:

$$\hat{b} = IDFT\,(DFT(\mathbf{b}) \oplus DFT\,(\tilde{X})). \qquad (11)$$

The model $\dot{m}_{p,j,k}$ of Eq-8 is then used with CMN on noisy speech. ~~Unfortunately, $\hat{b}$ is a function of both channel and background noise in all above cases. In other words, in presence of noise, there is no guarantee that the channel will be removed by such a vector, as is for CMN.~~

Change the paragraph that begins on page 8, line 17 to:
A ~~subset of WAVES~~ database containing ~~hands-free~~ recordings in a car was used. ~~, which consists of three recording sessions: parked trn (car parked, engine off), parked (car parked, engine off), and city-driving (car driven on a stop and go basis).~~

Remove the paragraph beginning on page 8, line 21 starting with "In each session" and ending with "10 dynamic coefficients".

Change the paragraph beginning on page 8, line 26 as follows:
HMMs used in all experiments ~~are~~ were trained using ~~in TIDIGITS~~ clean speech data. Utterance-based cepstral mean normalization ~~is~~ was used. ~~The HMMs contain 1957~~

~~mean vectors, and 270 diagonal variances. Evaluated on TIGIDIT test set, the recognizer gives 0.36% word error rate.~~

Remove the paragraph beginning on page 9, line 1 starting with " To improve" and ending with "JAC (joint compensation of additive noise and convolutive distortion)".

On page 9, remove the Table 1 beginning on line 6 and description on lines 9 and 10.

Change the paragraph that begins on page 9, line 12 to the end of the specification on page 10 at line 8 as follows:

~~Table 1shows that:~~

- ~~Compared to noise-free recognition (WER) (0.36%), without any compensation (BASELENE) the recognition performance degrades severely.~~
- ~~CMN effectively reduces the WER for parked data, but is not effective for driving conditions where additive noise becomes dominant.~~
- ~~PMC substantially reduces the WER for driving conditions, but gives poor results for parked data where microphone mismatch is dominant.~~
- ~~All JAC cases give lower WER than non-JAC methods.~~
- ~~Simplifying Eq-7 to Eq-9 then to Eq-10 results in progressive increase in WER, although the degradation is not severe. Especially, information in PDF probability is not critical to the performance.~~
- ~~Simplified JAC gives lowest WER in all tests.~~ Experimental results show that the new invented method ~~For this hands-free speech recognition, the new met~~ reduces word error rate by 61% ~~for parked condition and~~ to 94% relative to

baseline performance~~for~~ depending on ~~city~~ driving condition, and the new method is superior to other reported methods.